IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

BRITISH TELECOMMUNICATIONS PLC,	)	
	)	
Plaintiff,	)	
	)	
v.	)	C.A. No. 18-366 (WCB)
	)	
IAC/INTERACTIVECORP, MATCH	)	
GROUP, INC., MATCH GROUP, LLC and	)	
VIMEO, INC.,	)	
	)	
Defendants.	)	

DEFENDANTS' NOTICE OF SUBPOENA TO AMAZON.COM, INC.

PLEASE TAKE NOTICE that, pursuant to Rule 45 of the Federal Rules of Civil

Procedure, IAC/InterActiveCorp, Match Group, Inc. and Match Group, LLC (collectively,

"Defendants"), by and through their counsel, will serve the Subpoena attached as Exhibit A on

Amazon.com, Inc.

MORRIS, NICHOLS, ARSHT & TUNNELL LLP

/s/ Jack B. Blumenfeld

Jack B. Blumenfeld (#1014)
1201 North Market Street
P.O. Box 1347
Wilmington, DE 19899
(302) 658-9200
jblumenfeld@mnat.com

Attorneys for Defendants IAC/Interactive Corp, Match Group, Inc. and Match Group, LLC

OF COUNSEL:

Robert L. Hails
T. Cy Walker
BAKER & HOSTETLER LLP
Washington Square, Suite 1100
1050 Connecticut Avenue, N.W.
Washington, DC 20036-5403
(202) 861-1500

Fred Grasso
GRASSO PLLC
1818 Library Street, Suite 500
Reston, VA 20190
(571) 222-6520

2

Michael N. Zachary
BUNSOW DE MORY LLP
701 El Camino Real
Redwood City, CA 94063
(650) 351-7248

November 27, 2019

EXHIBIT A

UNITED STATES DISTRICT COURT

for the

District of Delaware

British Telecommunications PLC	)	
Plaintiff	)	
v.	)	Civil Action No. 18-cv-0366-WCB
IAC/InterActiveCorp, Match Group, Inc., Match Group, LLC, and Vimeo, Inc.	)	
Defendants	)	

SUBPOENA TO TESTIFY AT A DEPOSITION IN A CIVIL ACTION AND TO PRODUCE DOCUMENTS, INFORMATION, OR OBJECTS

To: Amazon.com, Inc.
c/o: Corporation Service Company 300 Deschutes Way SW, Suite 304 Tumwater, WA 98501

(Name of person to whom this subpoena is directed)

☒   *Testimony:* **YOU ARE COMMANDED** to appear at the time, date, and place set forth below to testify at a deposition to be taken in this civil action. If you are an organization, you must designate one or more officers, directors, or managing agents, or designate other persons who consent to testify on your behalf about the following matters, or those set forth in an attachment:

Place:	Baker & Hostetler LLP 999 Third Avenue Suite 3600 Seattle, WA 98104-4040	Date and Time: December 23, 2019 at 9 am pt

The deposition will be recorded by this method: Stenographically and video.

☒   *Production:* **YOU ARE COMMANDED** to produce at the time, date, and place set forth below the following documents, electronically stored information, or objects, and to permit inspection, copying, testing, or sampling of the material:

Place:	Baker & Hostetler LLP 999 Third Avenue Suite 3600 Seattle, WA 98104-4040	Date and Time: December 16, 2019 at 6 pm pt

The following provisions of Fed. R. Civ. P. 45 are attached – Rule 45(c), relating to the place of compliance; Rule 45(d), relating to your protection as a person subject to a subpoena; and Rule 45(e) and (g), relating to your duty to respond to this subpoena and the potential consequences of not doing so.

Date: 11/27/2019

CLERK OF COURT

	OR	/s/Michael N. Zachary
Signature of Clerk or Deputy Clerk		*Attorney's signature*

The name, address, e-mail address, and telephone number of the attorney representing *(name of party)*   Defendants

IAC/InterActiveCorp, Match Group, Inc, and Match Group, LLC   , who issues or requests this subpoena, are:

Michael Zachary, Bunsow De Mory LLP; 701 El Camino Rd, Redwood City, CA 94063; mzachary@bdiplaw.com ; (415) 426-4745

Notice to the person who issues or requests this subpoena

If this subpoena commands the production of documents, electronically stored information, or tangible things before trial, a notice and a copy of the subpoena must be served on each party in this case before it is served on the person to whom it is directed. Fed. R. Civ. P. 45(a)(4).

Civil Action Nos.   18-cv-0366-WCB

PROOF OF SERVICE

(This section should not be filed with the court unless required by Fed. R. Civ. P. 45.)

I received this subpoena for *(name of individual and title, if any)* _____

on *(date)* _____ .

☐   I served the subpoena by delivering a copy to the named individual as follows: _____

_____   on *(date)* _____ ; or

☐   I returned the subpoena unexecuted because: _____

_____ .

Unless the subpoena was issued on behalf of the United States, or one of its officers or agents, I have also tendered to the witness the fees for one day's attendance, and the mileage allowed by law, in the amount of

$ _____ .

My fees are $ _____ for travel and $ _____ for services, for a total of $ _____ .

I declare under penalty of perjury that this information is true.

Date: _____

_____
Server's signature

_____
Printed name and title

_____
Server's address

Additional information regarding attempted service, etc.:

4818-9873-3742.1

Federal Rule of Civil Procedure 45 (c), (d), (e), and (g) (Effective 12/1/13)

(c) Place of Compliance.

(1) *For a Trial, Hearing, or Deposition.* A subpoena may command a person to attend a trial, hearing, or deposition only as follows:
  (A) within 100 miles of where the person resides, is employed, or regularly transacts business in person; or
  (B) within the state where the person resides, is employed, or regularly transacts business in person, if the person
    (i) is a party or a party's officer; or
    (ii) is commanded to attend a trial and would not incur substantial expense.

(2) *For Other Discovery.* A subpoena may command:
  (A) production of documents, electronically stored information, or tangible things at a place within 100 miles of where the person resides, is employed, or regularly transacts business in person; and
  (B) inspection of premises at the premises to be inspected.

(d) Protecting a Person Subject to a Subpoena; Enforcement.

(1) *Avoiding Undue Burden or Expense; Sanctions.* A party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena. The court for the district where compliance is required must enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney's fees—on a party or attorney who fails to comply.

(2) *Command to Produce Materials or Permit Inspection.*
  (A) *Appearance Not Required.* A person commanded to produce documents, electronically stored information, or tangible things, or to permit the inspection of premises, need not appear in person at the place of production or inspection unless also commanded to appear for a deposition, hearing, or trial.
  (B) *Objections.* A person commanded to produce documents or tangible things or to permit inspection may serve on the party or attorney designated in the subpoena a written objection to inspecting, copying, testing, or sampling any or all of the materials or to inspecting the premises—or to producing electronically stored information in the form or forms requested. The objection must be served before the earlier of the time specified for compliance or 14 days after the subpoena is served. If an objection is made, the following rules apply:
    (i) At any time, on notice to the commanded person, the serving party may move the court for the district where compliance is required for an order compelling production or inspection.
    (ii) These acts may be required only as directed in the order, and the order must protect a person who is neither a party nor a party's officer from significant expense resulting from compliance.

(3) *Quashing or Modifying a Subpoena.*

  (A) *When Required.* On timely motion, the court for the district where compliance is required must quash or modify a subpoena that:

    (i) fails to allow a reasonable time to comply;
    (ii) requires a person to comply beyond the geographical limits specified in Rule 45(c);
    (iii) requires disclosure of privileged or other protected matter, if no exception or waiver applies; or
    (iv) subjects a person to undue burden.
  (B) *When Permitted.* To protect a person subject to or affected by a subpoena, the court for the district where compliance is required may, on motion, quash or modify the subpoena if it requires:

    (i) disclosing a trade secret or other confidential research, development, or commercial information; or
    (ii) disclosing an unretained expert's opinion or information that does not describe specific occurrences in dispute and results from the expert's study that was not requested by a party.
  (C) *Specifying Conditions as an Alternative.* In the circumstances described in Rule 45(d)(3)(B), the court may, instead of quashing or modifying a subpoena, order appearance or production under specified conditions if the serving party:
    (i) shows a substantial need for the testimony or material that cannot be otherwise met without undue hardship; and
    (ii) ensures that the subpoenaed person will be reasonably compensated.

(e) Duties in Responding to a Subpoena.

(1) *Producing Documents or Electronically Stored Information.* These procedures apply to producing documents or electronically stored information:
  (A) *Documents.* A person responding to a subpoena to produce documents must produce them as they are kept in the ordinary course of business or must organize and label them to correspond to the categories in the demand.
  (B) *Form for Producing Electronically Stored Information Not Specified.* If a subpoena does not specify a form for producing electronically stored information, the person responding must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms.
  (C) *Electronically Stored Information Produced in Only One Form.* The person responding need not produce the same electronically stored information in more than one form.
  (D) *Inaccessible Electronically Stored Information.* The person responding need not provide discovery of electronically stored information from sources that the person identifies as not reasonably accessible because of undue burden or cost. On motion to compel discovery or for a protective order, the person responding must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C). The court may specify conditions for the discovery.

(2) *Claiming Privilege or Protection.*
  (A) *Information Withheld.* A person withholding subpoenaed information under a claim that it is privileged or subject to protection as trial-preparation material must:
    (i) expressly make the claim; and
    (ii) describe the nature of the withheld documents, communications, or tangible things in a manner that, without revealing information itself privileged or protected, will enable the parties to assess the claim.
  (B) *Information Produced.* If information produced in response to a subpoena is subject to a claim of privilege or of protection as trial-preparation material, the person making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has; must not use or disclose the information until the claim is resolved; must take reasonable steps to retrieve the information if the party disclosed it before being notified; and may promptly present the information under seal to the court for the district where compliance is required for a determination of the claim. The person who produced the information must preserve the information until the claim is resolved.

(g) Contempt.
The court for the district where compliance is required—and also, after a motion is transferred, the issuing court—may hold in contempt a person who, having been served, fails without adequate excuse to obey the subpoena or an order related to it.

For access to subpoena materials, see Fed. R. Civ. P. 45(a) Committee Note (2013).

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

British Telecommunications PLC, Plaintiff, v. IAC/InterActiveCorp, Match Group, Inc., Match Group, LLC, and Vimeo, Inc., Defendants.	C.A. No.: 18-366-WCB

ATTACHMENTS TO THE SUBPOENA TO AMAZON.COM, INC.

Of Counsel:

Robert L. Hails (rhails@bakerlaw.com)
T. Cy Walker (cwalker@bakerlaw.com)
Cassandra J. Simmons
(csimmons@bakerlaw.com)
BAKER & HOSTETLER LLP
Washington Square, Suite 1100
1050 Connecticut Avenue, N.W.
Washington, DC 20036-5403

Fred Grasso (fgrasso@grassoip.com)
GRASSO, PLLC
1818 Library Street, Suite 500
Reston, VA 20190

Micheal Zachary (mzachary@bdiplaw.com)
BUNSOW DE MORY LLP
701 El Camino Real,
Redwood City, CA 94063

Dated: November 27, 2019

Jack B. Blumenfeld (#1014)
MORRIS, NICHOLS, ARSHT & TUNNELL LLP
1201 North Market Street
P.O. Box 1347
Wilmington, DE 19899
Tel.:   (302) 658-9200
Email: jblumenfeld@mnat.com

Counsel for IAC/InterActiveCorp, Match Group, Inc., Match Group, LLC, and Vimeo, Inc.

ATTACHMENT A

DEFINITIONS

1.   "Amazon," "you," or "your" means Amazon.com, Inc. its subsidiaries, divisions, predecessor and successor companies, affiliates, parents, any partnership or joint venture to which it is a party, and each of its employees, agents, officers, directors, and representatives, including any person who served in any of these capacities during any relevant time period.

2.   "BT" means British Telecommunications PLC, its subsidiaries, divisions, predecessor and successor companies, affiliates, parents, any partnership or joint venture to which it is a party, and each of its employees, agents, officers, directors, and representatives, including any person who served in any of these capacities during any relevant time period.

3.   "Recommendation Algorithms" means software algorithms used by Amazon.com between January 1, 2000 and December 31, 2002, to recommend items to online customers, including but not limited to the algorithms referenced in Linden, G., Smith, B., and York, J., *Amazon.com Recommendations: Item-to-Item Collaborative Filtering*, IEEE Internet Computing, 76-80 (Jan/Feb 2003) (Attachment D).

4.   "Document" is synonymous in meaning and equal in scope to its usage in Rule 34(a)(1)(A) of the Federal Rules of Civil Procedure. The term "document" refers to any document now or at any time your possession, custody, or control. A person is deemed in control of a document if the person has any ownership, possession, or custody of the document, or the right to secure the document or a copy thereof from any person or public or private entity having physical possession thereof.

5.   "ESI" refers to electronically stored information and is synonymous in meaning and equal in scope to its usage in Rule 34(a)(1)(A) of the Federal Rules of Civil Procedure. The

term "ESI" includes both active and residual ESI kept in the ordinary course of business. Residual

ESI includes, but is not limited to, deleted files, overwritten files, file fragments, or other data

found in ambient space on electronic storage media. Active ESI includes, but is not limited to,

information readily available and accessible to computer users through existing file management

programs.

      6.   "Person" means any natural person or any legal entity, including, any business or

governmental entity, organization, or association.

      7.   The terms "and" and "or" shall be construed either disjunctively or conjunctively

as necessary to bring within the scope of the discovery request all responses and documents that

might otherwise be construed to be outside its scope.

      8.   The use of the singular form of any word shall include the plural, and vice versa,

as necessary to bring within the scope of the discovery request all responses and documents that

might otherwise be construed to be outside its scope.

      9.   The term "including" shall mean including, but not limited to.

INSTRUCTIONS

      1.   The Production Requests (Attachment B) and Deposition Topics (Attachment C)

are intended to be narrow and impose minimal burden on you. If you have questions regarding the

requests or how to comply with them, please call Michael Zachary at (415) 426-4745.

      2.   The subpoena covers all documents and ESI in your possession, custody, or control.

      3.   All responsive documents and ESI should be produced as kept in the usual course

of business with any identifying folders, labels, file markings, or similar identifying features.

      4.   Responsive documents and ESI should be produced as native files or as text

searchable image files (e.g., PDF or TIFF) or in native format with metadata preserved.

5.   If you consider any responsive documents or ESI to contain confidential business information that you would like protected against public disclosure, you should mark and produce such documents or ESI as set forth in the governing Protective Order (Attachment E).

6.   If there are no responsive documents or ESI in response to a Production Request, you should state so in writing.

7.   If you object to any Production Request, you should produce all documents and ESI responsive to the non-objectionable portion of the Production Request.

8.   If you withhold any responsive documents or ESI based on an objection, you should state so in writing and give a general description of the documents or ESI being withheld.

9.   If you withhold any document or ESI pursuant to a claim of privilege, you should state the basis for the assertion of privilege in accordance with Federal Rule of Civil Procedure 26.

10.   If only a portion of a document or electronic file is responsive to a Production Request, you still should produce the document or electronic file in its entirety.

11.   If any document or ESI responsive to the Production Requests was but is no longer in your possession, custody, or control, you should identify the last known custodian(s) of such documents or ESI and the person(s) who made the decision to transfer or dispose of the document.

12.   You shall designate one or more officers, directors, or representatives to testify on your behalf regarding the Deposition Topics in Attachment C. The persons so designated must testify about information known or reasonably available to you, in accordance with Federal Rule of Civil Procedure 30(b)(6).

ATTACHMENT B

PRODUCTION REQUESTS

REQUEST NO. 1.   All non-privileged documents and ESI relating to any discussions, communications, or agreements you have had with BT over the past 8 years regarding actual or potential licensing or enforcement of any patent or patents owned by BT.

REQUEST NO. 2.   All non-privileged documents and ESI relating to any statements, summaries, assessments, analyses, reports, or opinions regarding the value of any patent or patents owned by BT, the value to you of a license to BT's patent(s), the amount of a royalty you would reasonably pay for use of BT's patent(s), your potential infringement of BT's patent(s), and/or the invalidity and/or unenforceability of BT's patents(s).

REQUEST NO. 3.   Documents and/or ESI sufficient to identify all Recommendation Algorithms used in the United States by Amazon.com between January 1, 2000 and December 31, 2002, and all inputs and outputs to such algorithms.

REQUEST NO. 4.   Documents and/or ESI sufficient to identify public uses of each Recommendation Algorithm identified in response to Request No. 3.

REQUEST NO. 5.   Documents, source code, and/or ESI sufficient to describe the function and operation of each Recommendation Algorithm identified in response to Request No. 3, including but not limited to the rules, algorithms, and instructions used to create, store, and update user-specific data, including user-profile data, user-preference data, and user-activity data.

1

ATTACHMENT C

DEPOSITION TOPICS

TOPIC NO. 1.   Your discussions, communications, and agreements with BT over the past 8 years regarding actual or potential licensing or enforcement of any patent or patents owned by BT and the related documents and ESI produced in response to Production Request No. 1.

TOPIC NO. 2.   Statements, summaries, assessments, analyses, reports, and opinions in your possession regarding the value of any patent or patents owned by BT, the value to you of a license to BT's patent(s), the amount of a royalty you would reasonably pay for use of BT's patent(s), your potential infringement of BT's patent(s), and/or the invalidity and/or unenforceability of BT's patent(s), and the related documents and ESI produced in response to Production Request No. 2.

TOPIC NO. 3.   Identification of the Recommendation Algorithms used in the United States by Amazon.com between January 1, 2000 and December 31, 2002, the inputs and outputs to such algorithms, and related documents and ESI produced in response to Production Request No. 3.

TOPIC NO. 4.   Identification of public uses of the Recommendation Algorithm responsive to Topic No. 3, and related documents and ESI produced in response to Production Request No. 4.

TOPIC NO. 5.   The function and operation of the Recommendation Algorithms referenced in Topic No. 3, including but not limited to the rules, algorithms, and instructions used to create, store, and update user-specific data, including user-profile data, user-preference data,

and user-activity data, and the identification of such rules, algorithms, and instructions in the related documents, source code, and ESI produced in response to Production Request No. 6.

   TOPIC NO. 7.   Authentication of the documents and ESI produced in response to this subpoena.

Industry Report

Amazon.com Recommendations

Item-to-Item Collaborative Filtering

Greg Linden, Brent Smith, and Jeremy York • *Amazon.com*

R ecommendation algorithms are best known for their use on e-commerce Web sites,[1] where they use input about a customer's interests to generate a list of recommended items. Many applications use only the items that customers purchase and explicitly rate to represent their interests, but they can also use other attributes, including items viewed, demographic data, subject interests, and favorite artists.

At Amazon.com, we use recommendation algorithms to personalize the online store for each customer. The store radically changes based on customer interests, showing programming titles to a software engineer and baby toys to a new mother. The click-through and conversion rates — two important measures of Web-based and email advertising effectiveness — vastly exceed those of untargeted content such as banner advertisements and top-seller lists.

E-commerce recommendation algorithms often operate in a challenging environment. For example:

- A large retailer might have huge amounts of data, tens of millions of customers and millions of distinct catalog items.
- Many applications require the results set to be returned in realtime, in no more than half a second, while still producing high-quality recommendations.
- New customers typically have extremely limited information, based on only a few purchases or product ratings.
- Older customers can have a glut of information, based on thousands of purchases and ratings.
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithm must respond immediately to new information.

There are three common approaches to solving the recommendation problem: traditional collaborative filtering, cluster models, and search-based methods. Here, we compare these methods with our algorithm, which we call *item-to-item collaborative filtering*. Unlike traditional collaborative filtering, our algorithm's online computation scales independently of the number of customers and number of items in the product catalog. Our algorithm produces recommendations in realtime, scales to massive data sets, and generates high-quality recommendations.

Recommendation Algorithms

Most recommendation algorithms start by finding a set of customers whose purchased and rated items overlap the user's purchased and rated items.[2] The algorithm aggregates items from these similar customers, eliminates items the user has already purchased or rated, and recommends the remaining items to the user. Two popular versions of these algorithms are *collaborative filtering* and *cluster models*. Other algorithms — including search-based methods and our own item-to-item collaborative filtering — focus on finding similar items, not similar customers. For each of the user's purchased and rated items, the algorithm attempts to find similar items. It then aggregates the similar items and recommends them.

Traditional Collaborative Filtering

A traditional collaborative filtering algorithm represents a customer as an N-dimensional vector of items, where N is the number of distinct catalog items. The components of the vector are positive for purchased or positively rated items and negative for negatively rated items. To compensate for

best-selling items, the algorithm typically multiplies the vector components by the inverse frequency (the inverse of the number of customers who have purchased or rated the item), making less well-known items much more relevant.[3] For almost all customers, this vector is extremely sparse.

The algorithm generates recommendations based on a few customers who are most similar to the user. It can measure the similarity of two customers, A and B, in various ways; a common method is to measure the cosine of the angle between the two vectors:[4]

$$similarity(\vec{A}, \vec{B}) = \cos(\vec{A}, \vec{B}) = \frac{\vec{A} \bullet \vec{B}}{\|\vec{A}\| * \|\vec{B}\|}$$

The algorithm can select recommendations from the similar customers' items using various methods as well, a common technique is to rank each item according to how many similar customers purchased it.

Using collaborative filtering to generate recommendations is computationally expensive. It is $O(MN)$ in the worst case, where M is the number of customers and N is the number of product catalog items, since it examines M customers and up to N items for each customer. However, because the average customer vector is extremely sparse, the algorithm's performance tends to be closer to $O(M + N)$. Scanning every customer is approximately $O(M)$, not $O(MN)$, because almost all customer vectors contain a small number of items, regardless of the size of the catalog. But there are a few customers who have purchased or rated a significant percentage of the catalog, requiring $O(N)$ processing time. Thus, the final performance of the algorithm is approximately $O(M + N)$. Even so, for very large data sets — such as 10 million or more customers and 1 million or more catalog items — the algorithm encounters severe performance and scaling issues.

It is possible to partially address these scaling issues by reducing the data size.[4] We can reduce M by randomly sampling the customers or discarding customers with few purchases, and reduce N by discarding very popular or unpopular items. It is also possible to reduce the number of items examined by a small, constant factor by partitioning the item space based on product category or subject classification. Dimensionality reduction techniques such as clustering and principal component analysis can reduce M or N by a large factor.[5]

Unfortunately, all these methods also reduce recommendation quality in several ways. First, if the algorithm examines only a small customer sample, the selected customers will be less similar to the user. Second, item-space partitioning restricts recommendations to a specific product or subject area. Third, if the algorithm discards the most popular or unpopular items, they will never appear as recommendations, and customers who have purchased only those items will not get recommendations. Dimensionality reduction techniques applied to the item space tend to have the same effect by eliminating low-frequency items. Dimensionality reduction applied to the customer space effectively groups similar customers into clusters; as we now describe, such clustering can also degrade recommendation quality.

Cluster Models

To find customers who are similar to the user, cluster models divide the customer base into many segments and treat the task as a classification problem. The algorithm's goal is to assign the user to the segment containing the most similar customers. It then uses the purchases and ratings of the customers in the segment to generate recommendations.

The segments typically are created using a clustering or other unsupervised learning algorithm, although some applications use manually determined segments. Using a similarity metric, a clustering algorithm groups the most similar customers together to form clusters or segments. Because optimal clustering over large data sets is impractical, most applications use various forms of greedy cluster generation. These algorithms typically start with an initial set of segments, which often contain one randomly selected customer each. They then repeatedly match customers to the existing segments, usually with some provision for creating new or merging existing segments.[6] For very large data sets — especially those with high dimensionality — sampling or dimensionality reduction is also necessary.

Once the algorithm generates the segments, it computes the user's similarity to vectors that summarize each segment, then chooses the segment with the strongest similarity and classifies the user accordingly. Some algorithms classify users into multiple segments and describe the strength of each relationship.[7]

Cluster models have better online scalability and performance than collaborative filtering[3] because they compare the user to a controlled number of segments rather than the entire cus-
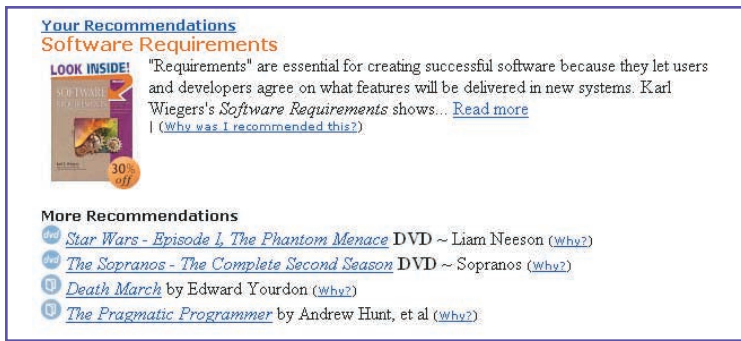
Industry Report

Figure 1. The "Your Recommendations" feature on the Amazon.com homepage. Using this feature, customers can sort recommendations and add their own product ratings.

Figure 2. Amazon.com shopping cart recommendations. The recommendations are based on the items in the customer's cart: The Pragmatic Programmer *and* Physics for Game Developers.

tomer base. The complex and expensive clustering computation is run offline. However, recommendation quality is low.[1] Cluster models group numerous customers together in a segment, match a user to a segment, and then consider all customers in the segment similar customers for the purpose of making recommendations. Because the similar customers that the cluster models find are not the most similar customers, the recommendations they produce are less relevant. It is possible to improve quality by using numerous fine-grained segments, but then online user–segment classification becomes almost as expensive as finding similar customers using collaborative filtering.

Search-Based Methods
Search- or content-based methods treat the recommendations problem as a search for related items.[8] Given the user's purchased and rated items, the algorithm constructs a search query to find other popular items by the same author, artist, or director, or with similar keywords or subjects. If a customer buys the Godfather DVD Collection, for example, the system might recommend other crime drama titles, other titles starring Marlon Brando, or other movies directed by Francis Ford Coppola.

If the user has few purchases or ratings, search-based recommendation algorithms scale and perform well. For users with thousands of purchases, however, it's impractical to base a query on all the items. The algorithm must use a subset or summary of the data, reducing quality. In all cases, recommendation quality is relatively poor. The recommendations are often either too general (such as best-selling drama DVD titles) or too narrow (such as all books by the same author). Recommendations should help a customer find and discover new, relevant, and interesting items. Popular items by the same author or in the same subject category fail to achieve this goal.

Item-to-Item Collaborative Filtering
Amazon.com uses recommendations as a targeted marketing tool in many email campaigns and on most of its Web sites' pages, including the high-traffic Amazon.com homepage. Clicking on the "Your Recommendations" link leads customers to an area where they can filter their recommendations by product line and subject area, rate the recommended products, rate their previous purchases, and see why items are recommended (see Figure 1).

As Figure 2 shows, our shopping cart recommendations, which offer customers product suggestions based on the items in their shopping cart. The feature is similar to the impulse items in a supermarket checkout line, but our impulse items are targeted to each customer.

Amazon.com extensively uses recommendation algorithms to personalize its Web site to each customer's interests. Because existing recommendation algorithms cannot scale to Amazon.com's tens of millions of customers and products, we developed our own. Our algorithm, item-to-item collaborative filtering, scales to massive data sets and produces high-quality recommendations in real time.

How It Works
Rather than matching the user to similar customers, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list.[9]

To determine the most-similar match for a given item, the algorithm builds a similar-items table by finding items that customers tend to purchase together. We could build a product-to-product matrix by iterating through all item pairs and computing a similarity metric for each pair. However, many product pairs have no common customers, and thus the approach is inefficient in terms of processing time and memory usage. The following

iterative algorithm provides a better approach by calculating the similarity between a single product and all related products:

```
For each item in product catalog, I₁
  For each customer C who purchased I₁
    For each item I₂ purchased by
       customer C
      Record that a customer purchased I₁
      and I₂
  For each item I₂
    Compute the similarity between I₁ and I₂
```

It's possible to compute the similarity between two items in various ways, but a common method is to use the cosine measure we described earlier, in which each vector corresponds to an item rather than a customer, and the vector's M dimensions correspond to customers who have purchased that item.

This offline computation of the similar-items table is extremely time intensive, with $O(N^2M)$ as worst case. In practice, however, it's closer to $O(NM)$, as most customers have very few purchases. Sampling customers who purchase best-selling titles reduces runtime even further, with little reduction in quality.

Given a similar-items table, the algorithm finds items similar to each of the user's purchases and ratings, aggregates those items, and then recommends the most popular or correlated items. This computation is very quick, depending only on the number of items the user purchased or rated.

Scalability: A Comparison
Amazon.com has more than 29 million customers and several million catalog items. Other major retailers have comparably large data sources. While all this data offers opportunity, it's also a curse, breaking the backs of algorithms designed for data sets three orders of magnitude smaller. Almost all existing algorithms were evaluated over small data sets. For example, the MovieLens data set[4] contains 35,000 customers and 3,000 items, and the EachMovie data set[3] contains 4,000 customers and 1,600 items.

For very large data sets, a scalable recommendation algorithm must perform the most expensive calculations offline. As a brief comparison shows, existing methods fall short:

- Traditional collaborative filtering does little or no offline computation, and its online computation scales with the number of customers and catalog items. The algorithm is impractical on large data sets, unless it uses dimensionality reduction, sampling, or partitioning — all of which reduce recommendation quality.
- Cluster models can perform much of the computation offline, but recommendation quality is relatively poor. To improve it, it's possible to increase the number of segments, but this makes the online user–segment classification expensive.
- Search-based models build keyword, category, and author indexes offline, but fail to provide recommendations with interesting, targeted titles. They also scale poorly for customers with numerous purchases and ratings.

The key to item-to-item collaborative filtering's scalability and performance is that it creates the expensive similar-items table offline. The algorithm's online component — looking up similar items for the user's purchases and ratings — scales independently of the catalog size or the total number of customers; it is dependent only on how many titles the user has purchased or rated. Thus, the algorithm is fast even for extremely large data sets. Because the algorithm recommends highly correlated similar items, recommendation quality is excellent.[10] Unlike traditional collaborative filtering, the algorithm also performs well with limited user data, producing high-quality recommendations based on as few as two or three items.

Conclusion
Recommendation algorithms provide an effective form of targeted marketing by creating a personalized shopping experience for each customer. For large retailers like Amazon.com, a good recommendation algorithm is scalable over very large customer bases and product catalogs, requires only subsecond processing time to generate online recommendations, is able to react immediately to changes in a user's data, and makes compelling recommendations for all users regardless of the number of purchases and ratings. Unlike other algorithms, item-to-item collaborative filtering is able to meet this challenge.

In the future, we expect the retail industry to more broadly apply recommendation algorithms for targeted marketing, both online and offline. While e-commerce businesses have the easiest vehicles for personalization, the technology's increased conversion rates as compared with traditional broad-scale approaches will also make it compelling to offline retailers for use in postal mailings, coupons, and other forms of customer communication.

References

1. J.B. Schafer, J.A. Konstan, and J. Reidl, "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, Kluwer Academic, 2001, pp. 115-153.
2. P. Resnick et al., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. ACM 1994 Conf. Computer Supported Cooperative Work*, ACM Press, 1994, pp. 175-186.
3. J. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proc. 14th Conf. Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1998, pp. 43-52.
4. B.M. Sarwarm et al., "Analysis of Recommendation Algorithms for E-Commerce," *ACM Conf. Electronic Commerce*, ACM Press, 2000, pp.158-167.
5. K. Goldberg et al., "Eigentaste: A Constant Time Collaborative Filtering Algorithm," *Information Retrieval J.*, vol. 4, no. 2, July 2001, pp. 133-151.
6. P.S. Bradley, U.M. Fayyad, and C. Reina, "Scaling Clustering Algorithms to Large Databases," *Knowledge Discovery and Data Mining*, Kluwer Academic, 1998, pp. 9-15.
7. L. Ungar and D. Foster, "Clustering Methods for Collaborative Filtering," *Proc. Workshop on Recommendation Systems*, AAAI Press, 1998.
8. M. Balabanovic and Y. Shoham, "Content-Based Collaborative Recommendation," *Comm. ACM*, Mar. 1997, pp. 66-72.
9. G.D. Linden, J.A. Jacobi, and E.A. Benson, *Collaborative Recommendations Using Item-to-Item Similarity Mappings*, US Patent 6,266,649 (to Amazon.com), Patent and Trademark Office, Washington, D.C., 2001.
10. B.M. Sarwar et al., "Item-Based Collaborative Filtering Recommendation Algorithms," *10th Int'l World Wide Web Conference*, ACM Press, 2001, pp. 285-295.

Greg Linden was cofounder, researcher, and senior manager in the Amazon.com Personalization Group, where he designed and developed the recommendation algorithm. He is currently a graduate student in management in the Sloan Program at Stanford University's Graduate School of Business. His research interests include recommendation systems, personalization, data mining, and artificial intelligence. Linden received an MS in computer science from the University of Washington. Contact him at Linden_Greg@gsb.stanford.edu.

Brent Smith leads the Automated Merchandising team at Amazon.com. His research interests include data mining, machine learning, and recommendation systems. He received a BS in mathematics from the University of California, San Diego, and an MS in mathematics from the University of Washington, where he did graduate work in differential geometry. Contact him at smithbr@amazon.com.

Jeremy York leads the Automated Content Selection and Delivery team at Amazon.com. His interests include statistical models for categorical data, recommendation systems, and optimal choice of Web site display components. He received a PhD in statistics from the University of Washington, where his thesis won the Leonard J. Savage award for best thesis in applied Bayesian econometrics and statistics. Contact him at jeremy@amazon.com.

CERTIFICATE OF SERVICE

I hereby certify that on November 27, 2019, I caused the foregoing to be electronically

filed with the Clerk of the Court using CM/ECF, which will send notification of such filing to all

registered participants.

I further certify that I caused copies of the foregoing document to be served on

November 27, 2019, upon the following in the manner indicated:

Philip A. Rovner, Esquire   *VIA ELECTRONIC MAIL*
Jonathan A. Choa, Esquire
POTTER ANDERSON & CORROON LLP
1313 North Market Street
Hercules Plaza, 6th Floor
Wilmington, DE 19801
Attorneys for Plaintiff

Daniel A. Boehnen, Esquire   *VIA ELECTRONIC MAIL*
Grantland G. Drutchas, Esquire
Jeffrey P. Armstrong, Esquire
George T. Lyons, III, Esquire
MCDONNELL BOEHNEN HULBERT
& BERGHOFF LLP
300 South Wacker Drive
Chicago, IL 60606
Attorneys for Plaintiff

/s/ *Jack B. Blumenfeld*

_____
Jack B. Blumenfeld (#1014)